ASCI I/O: End-to-End

End-to-End I/O Perfomance

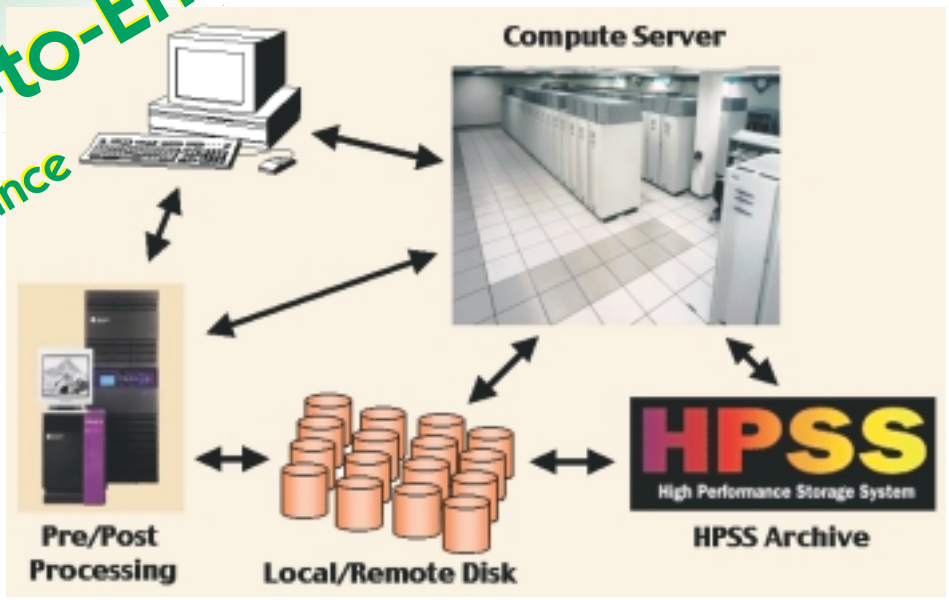

Figure 1. High-performance data access for application users.

Scalable Input/Output (I/O) for ASCI applications is an indispensable part of the infrastructure needed to support the high-level ASCI goals in weapons science, theory, and experimentation. Unless I/O transfer rates are in balance with the increasing computation speeds, ASCI applications will spend far too much time storing and retrieving data. To provide a long-term solution for Scalable-I/O it must be viewed in the end-to-end context of the several layers that must work together to deliver convenient and high-performance data access for application users, as shown in Figure 1.

Historical Perspective

Historically, scientists and engineers have depended on experimental results to validate their theories and designs. Today, experimentation is increasingly being replaced by computer modeling and simulation. Increased reliance on high-resolution simulation forces the scope and complexity of those applications to increase dramatically. Computational resources have kept up with the increased demand, but I/O subsystems have not.

Long term trends in computing show that processor speed and capacity is growing at a much higher rate than I/O transfer rates. The corresponding imbalance between I/O and compute power will only get worse with the introduction of larger clusters of faster processors. I/O "bottlenecks" slow modeling and simulation application output, as well as visualization application input. The I/O "bottleneck" appears at different levels of software/hardware abstractions, e.g., in data format, application read/write, file systems, and disk.

End-to-End I/O Layers

Between applications and file systems, ASCI has defined three primary layers in the end-to-end I/O problem as shown in Figure 2. The Application Program Interface Layer provides an interface for ASCI applications. The Middle Layer provides data management capabilities. The Lower-Level I/O Interface provides I/O libraries for more efficient data handling.

It is estimated that ASCI applications will require 1–2 GB/sec I/O rates per TeraOPS. The requirements are estimated for single-application delivered data rates for parallel disk systems on ASCI computers. Significant I/O parallelism is necessary to satisfy such requirements. ASCI must develop flexible operating systems, I/O libraries, and database and information management systems, that can match such large-scale application I/O demands to storage hardware realities and visualization needs. An ASCI tri-lab study group for data visualization recommended an integrated systems approach and solution. Such a system must include the end-to-end flow of data from generation to storage and must support data retrieval, visualization, data mining, and sophisticated presentation capabilities.

Data Access Layers

The end-to-end data access layers shown in Figure 3 include the full range of software and hardware infrastructure needed. The layers of primary interest to Scalable-I/O are from the Scalable Access Layer on down. The layers above may be viewed as client applications by Scalable-I/O. The requirements are to deliver high-performance I/O not only on the systems of today, but also that scales up to the 10, 30, and 100 TeraOPS

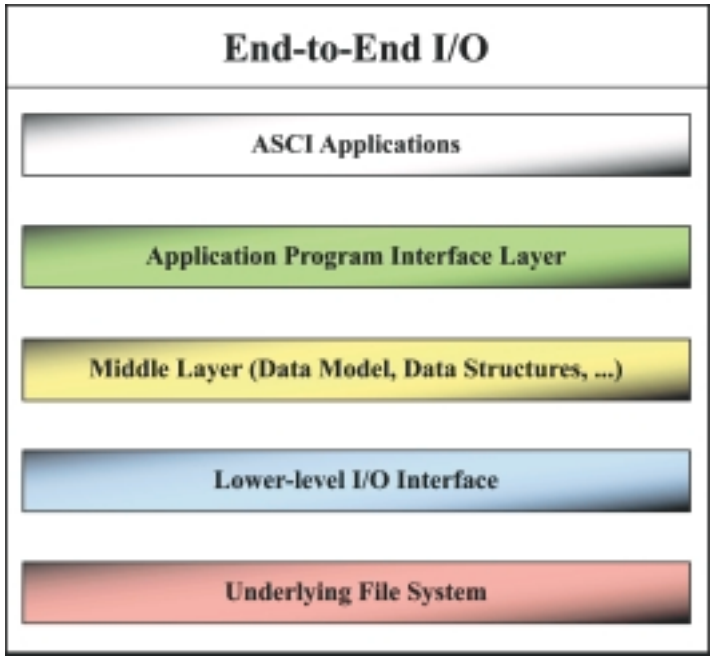

Figure 2. End-to-End I/O.

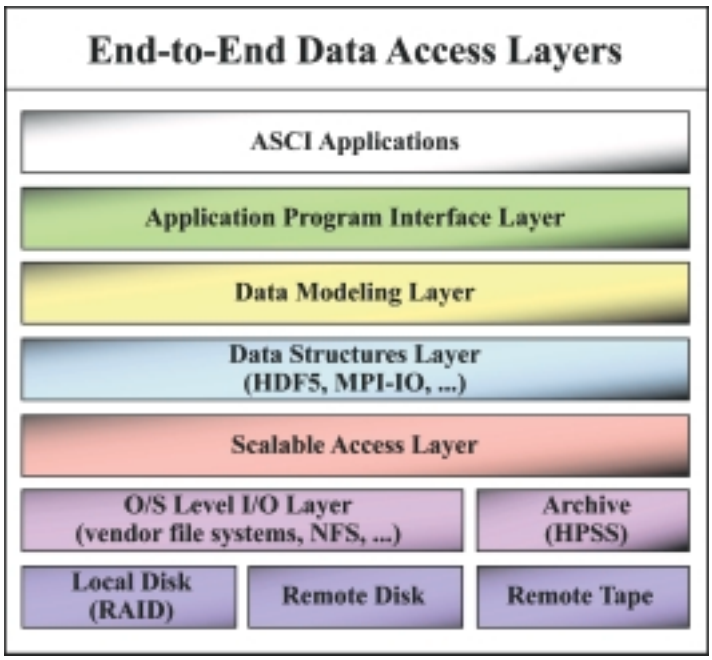

Figure 3. End-to-End Data Access Layers.

machines that are coming. This capability must be provided without requiring extensive rewrites of the client layers and without the time and effort to re-develop the software for the Scalable-I/O layers.

Among the three labs, some consensus has developed in the area of data models and formats using the HDF5 library as a portability layer. For scalable and parallel I/O, the labs have chosen MPI-IO as a software application programming interface standard because of its portability. But optimum, or even adequate, performance of MPI-IO depends on the underlying file system implementation. The functionality of MPI-IO may also be limited by existing file system restrictions to global access. In the scenario of end-to-end I/O, the file system is the lowest level software link to the disk subsystem visible from an application.

Summary

I/O and data handling remain fundamental bottlenecks in high-performance computing. The requirement for ASCI Scalable I/O is to deliver very high transfer rates for the ASCI applications in a reliable way that scales up with the computing capability of each new system. The approach must consider I/O in its end-to-end context.

For more information on the ASCI Scalable I/O project, please contact:

Judy Sturtevant, (505)845-9448, jesturt@sandia.gov;

Tyce McLarty, (505) 667-6034, ttm@lanl.gov; or

Kim Yates, (925) 423-5535, rkyates@llnl.gov

Work performed under the auspices of the U.S. DOE by LLNL under contract No. W-7405-ENG-48